

The background of the slide is a high-resolution image of Earth as seen from space. It shows a curved horizon with a deep blue sky above and a detailed view of the Earth's surface below, including mountain ranges, valleys, and some snow-covered areas. A small orange horizontal line is positioned above the title text.

# Multivariate Modelling for Hardware Estimates in Predevelopment Projects – OHV expanding toolset

Topic: Cost analysis and modelling

*Giulia Cambone – Subcontractor Manager*

*Dr. Luca Spani Molella – Project Procurement Manager*

# AGENDA

---



- 1. Introduction**
- 2. Parametric cost modelling theory concepts**
- 3. OHV Cost Estimating Relationships**
- 4. Conclusion**

A short, solid orange horizontal bar is positioned above the title.

# INTRODUCTION



# Why the need for Multivariate Modelling in early phase projects

The need for estimating the cost of non-Off-The-Shelf (OTS) items

## Early phase estimates

*Limited information, decisions needed*

### Portfolio-based Hardware



- ✓ Existing portfolio
- ✓ Heritage pricing
- ✓ Analogy estimates



### Tailoring-based Hardware



- ✗ No catalogue items
- ✗ Mission-specific design
- ✗ Limited cost visibility

### Multivariate parametric cost modelling

In early project phases, cost estimates are needed before the hardware design exists. Multivariate modelling bridges that gap.

# Tool purpose and applicability

Providing *what if* scenarios & supporting trade-offs

## Purpose: Support early-phase decision making

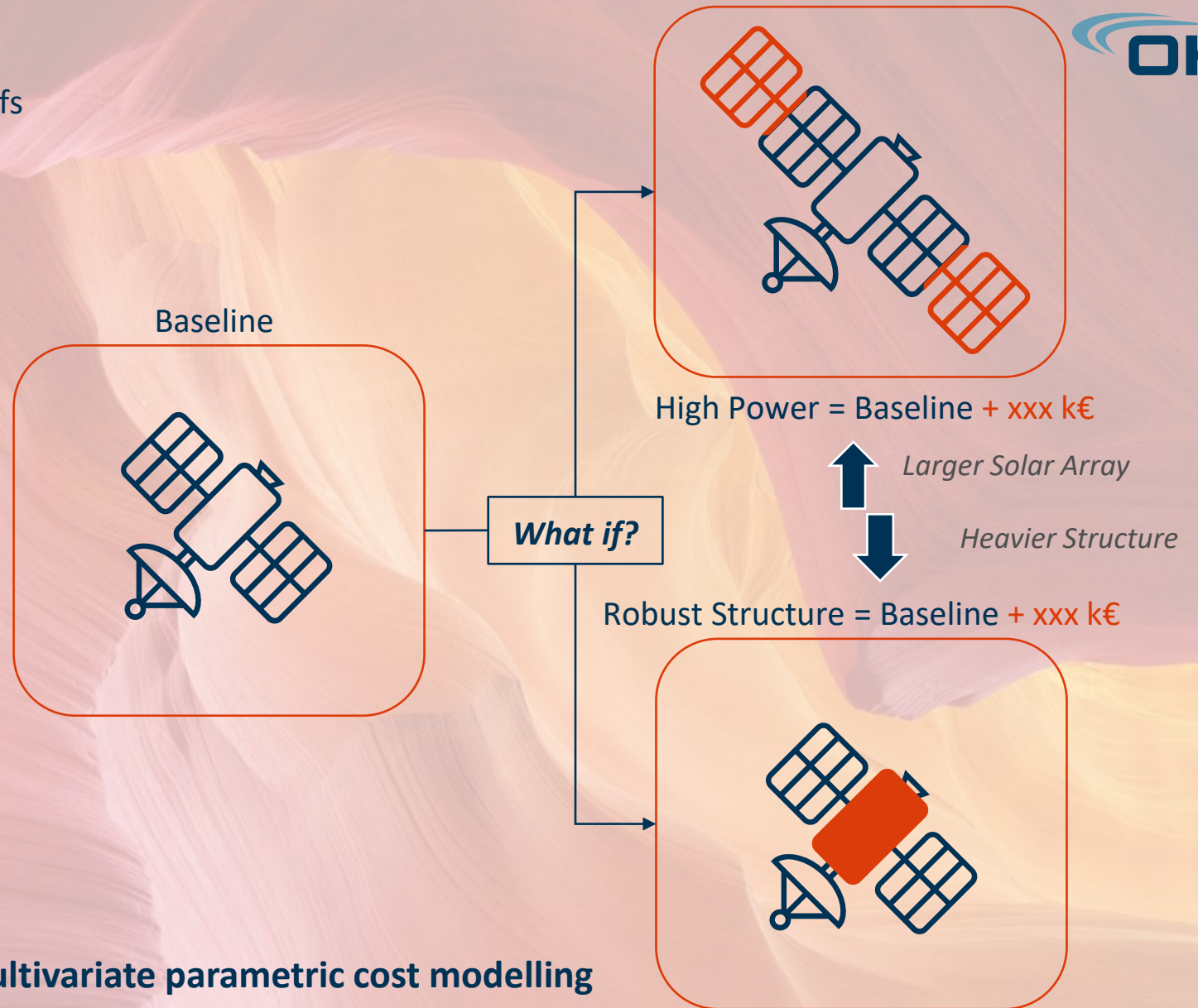
- Generate rapid hardware cost estimates
- Quantify cost sensitivity to key design parameters
- Compare alternative design concepts
- Provide consistent and repeatable estimates

## Applicability: Ideal for Phase 0/A studies

- Limited technical definition
- Frequent requirement changes
- Multiple architecture candidates
- Need for rapid cost feedback

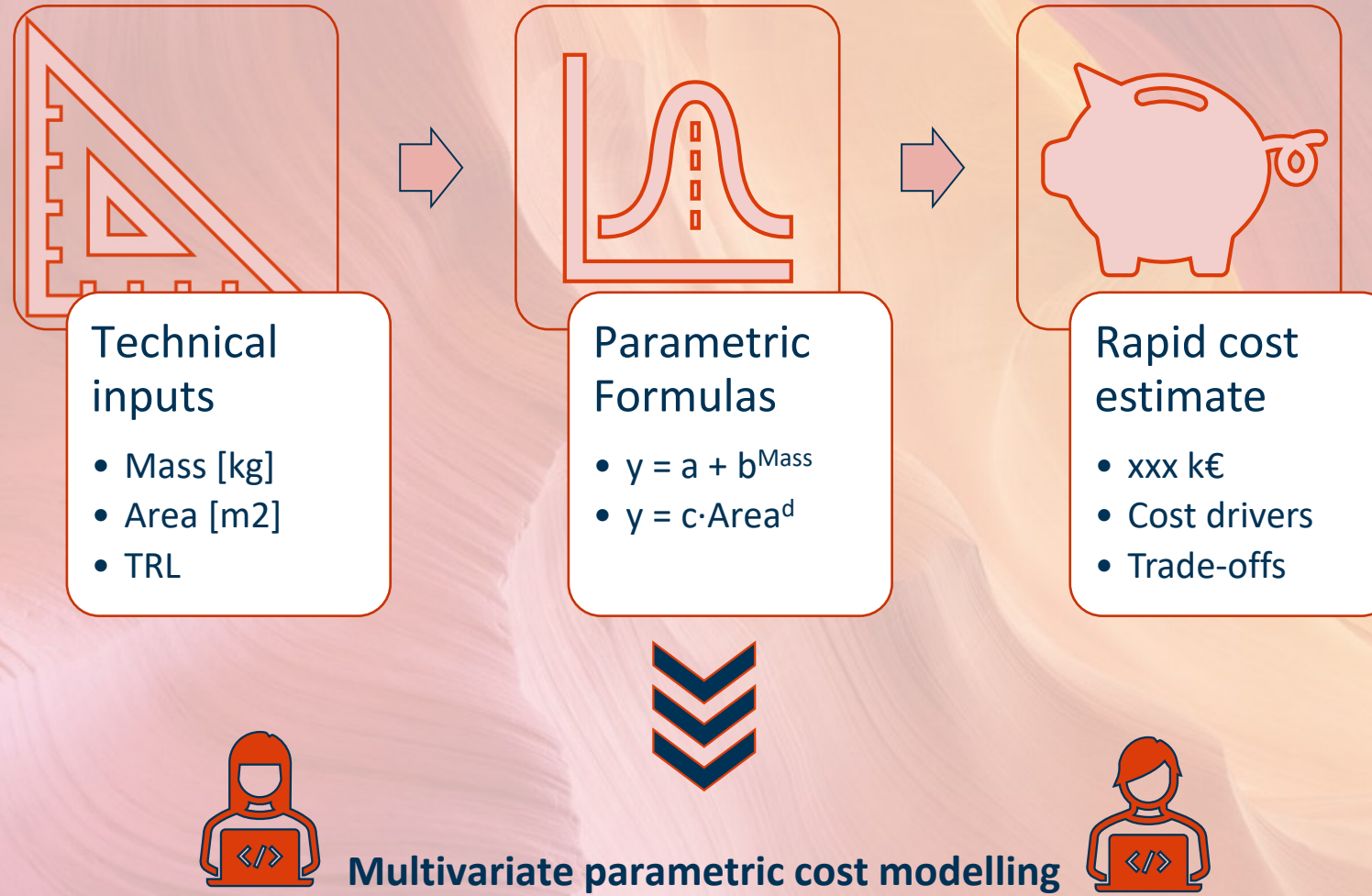
## Multivariate parametric cost modelling

Explore design alternatives and understand their cost impact before detailed design is available.



# Tool purpose and applicability

Early phase projects: rapid estimates from top-level inputs



The model supports the estimator; it does not replace the estimator.

# THEORY CONCEPTS

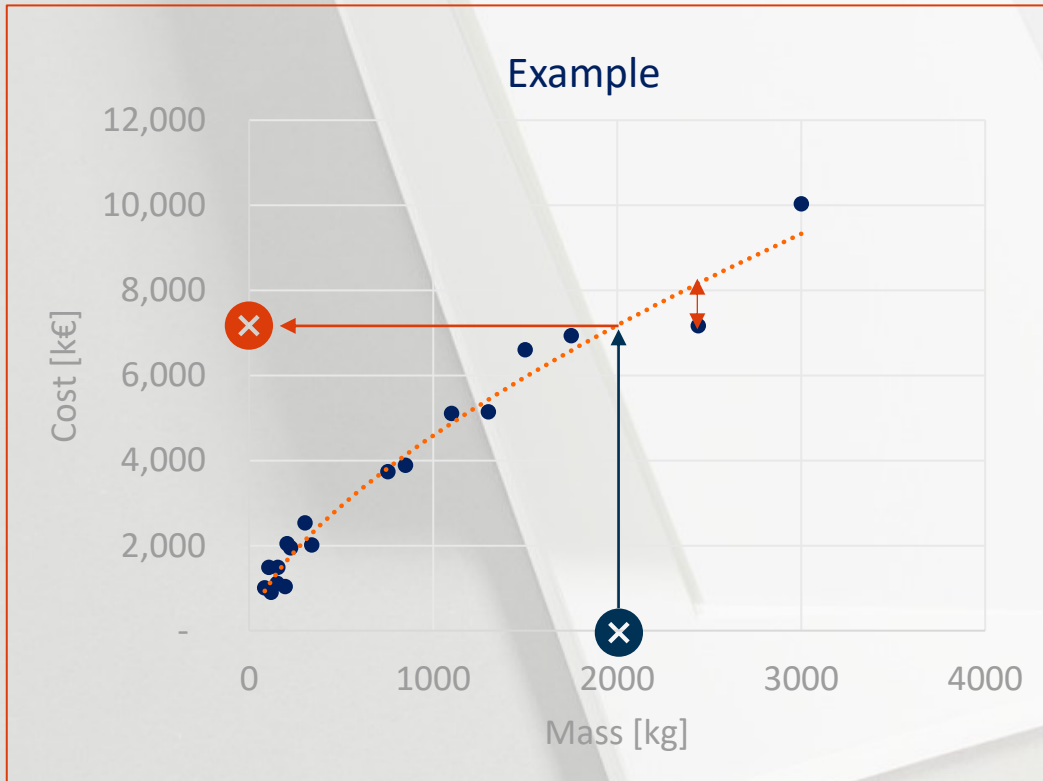


# How are CERs derived

## Building the formula

### ■ CERs | Cost Estimating Relationships

- a mathematical formula where a dependent variable (cost) is expressed as a function of one (single-variate) or more (multi-variate) independent cost-driving parameters (e.g. mass, power)



$a$  = intercept,  $b$  = characteristic of the product,  $c$  = lower k€/kg as Mass increases

### ■ How are CERs derived?

1. Make a hypothesis on the main Cost Driver (based on what makes physically sense, what is driving the performance): e.g. Mass
2. Observe if there is a correlation between the two (positive correlation in example shown)
3. Decide which functional form to use (e.g. linear  $y = a + b \cdot x$ , or other): in the example here, we choose:
  - $y = a + b \cdot x^c \rightarrow \text{Estimated Cost} = a + b \cdot \text{Mass}^c$
4. Gather Cost and Mass from same family products – **historical database**
5. Using  $x, y$  (**Mass, Real Cost**) as inputs, calculate for which parameters  $a, b, c$  the curve is better fitting the points in the scatter chart
  - **Mass, Real Cost**  $\rightarrow$  ZMPE method  $\rightarrow$  CER parameter  $a, b, c$
  - The ZMPE method\* minimizes the error between the **Real Cost** and the **Estimated Cost** (the distance between all the dots in the scatter chart and the curve  $\updownarrow$ )
    - **Estimated Cost** =  $a + b \cdot \text{Mass}^c$

\*see next slide

# How are CERs derived

## ZMPE Method

- **Cost** =  $a + b \cdot X_1^{c1} \cdot X_2^{c2} \cdot \dots \cdot X_m^{cm} \cdot d_1^{x_{m+1}} \cdot d_2^{x_{m+2}} \cdot \dots \cdot d_n^{x_n}$ 
  - CER formula
  - $X_i$  Quantitative Variables ,  $x_j$  Categorical Variables → Selected by the CER creator, based on: logic, physics, availability
  - $a, b, c_k, d_z$  CER parameters → ZMPE Method
  
- Calculation of CER parameters by a **constrained minimization process - Zero Percentage Bias Minimum Percentage Error (ZMPE)**:
  - Minimize the deviation of the estimated values from the actual values (the average error in the prediction)
  - While subject to the constraint of not consistently over/underestimating the actual values (residuals normally distributed)

### 1. Minimum Percentage Error

Minimize SSE (Sum of Squared Errors)  $\sum_i \left( \frac{y_i - \hat{y}}{\hat{y}} \right)^2$

### 2. Minimum Percentage Bias

Subject to constraint  $\text{BIAS} = \sum_i \left( \frac{y_i - \hat{y}}{\hat{y}} \right) = 0$

$y_i$  = actual values

$\hat{y}$  = estimated values =  $a + b \cdot X_1^{c1} \cdot X_2^{c2} \cdot \dots$

# Assessing CER Quality

## Statistical metrics

### SSE = Prediction error

- Measures how much of the estimate values deviate from actual values → Lower is better
- Objective:  $Min \sum_i \left( \frac{y_i - \hat{y}}{\hat{y}} \right)^2$

### BIAS = Systematic over / under estimation

- Measures whether the CER consistently overestimates or underestimates costs → Close to zero is better
- Objective:  $BIAS = \sum_i \left( \frac{y_i - \hat{y}}{\hat{y}} \right) = 0$

### R<sup>2</sup> = Goodness of fit

- Measures how much of the observed cost variability is explained by the CER → Close to one is better
- Objective:  $R^2 = 1 - \frac{(y_i - \hat{y})^2}{(y_i - \langle y \rangle)^2} \sim 1$

$y_i$  = actual values

$\hat{y}$  = estimated values =  $a + b \cdot X_1^{c1} \cdot X_2^{c2} \cdot \dots$

**We are not looking for correlations. We are looking for meaningful correlations.**

# OHB COST ESTIMATING RELATIONSHIPS



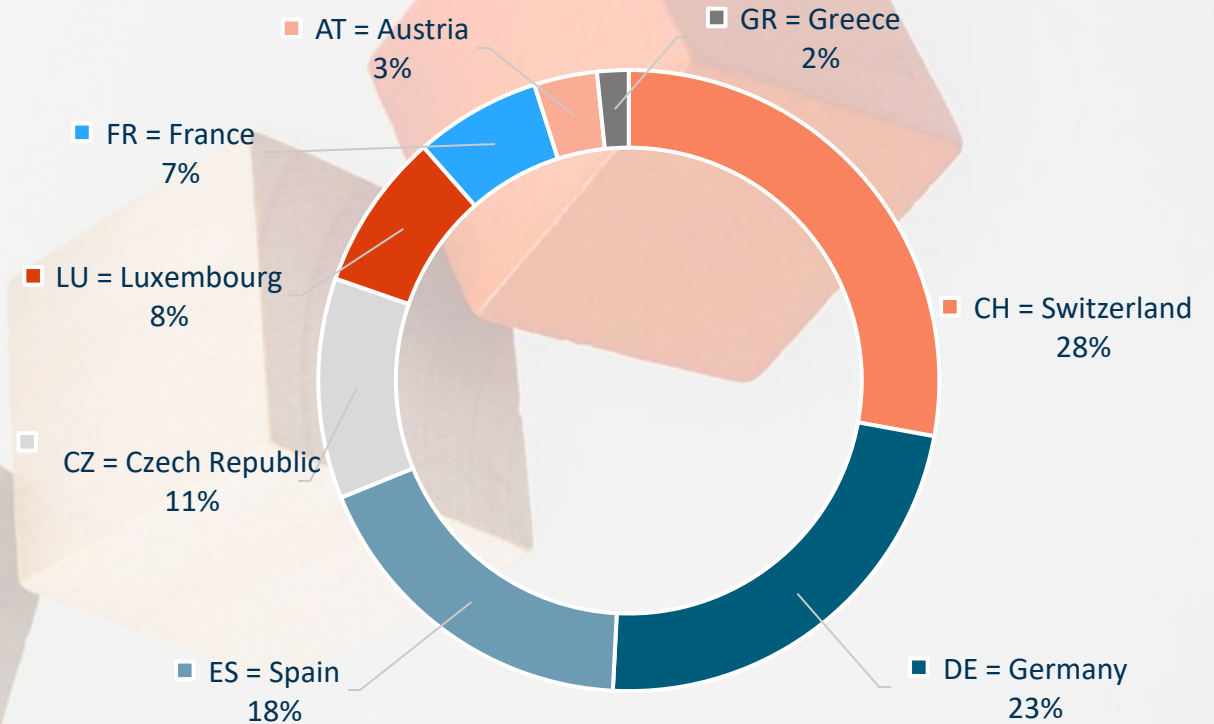
# CER - Structural Panels

## Scope of the dataset



- ✓ Costs normalized to same Economic Conditions (EC)
- ✓ RFIs (non-binding) and RFP (binding) responses included
- ✓ Earth Observation, Science and Connectivity programmes
- ✓ Common ESA procurement framework
- ✓ Subset of homogeneous proposals within OHb available data

## Proposals from each Country



**A diverse dataset, but a consistent procurement context**

# CER - Structural Panels

Cost drivers' selection rationale & Formula

 **Volume**

Hardware size

 **Technology Readiness**

Development effort

 **Material**

Manufacturing complexity

 **k - factor**

Contractual scope

$$\text{Structural Panel [k€]} = a + b \cdot \text{Volume [m}^3\text{]}^c \cdot d_{i=[1,2,3]} \cdot e_{j=[1,2,3]} \cdot k$$

- $a, b, c$  = CER parameters
- Volume [m<sup>3</sup>] = main Quantitative variable
- $d$  = Technology Readiness,  $e$  = Material
  - $d_1$  = Recurrent Design,  $d_2$  = Modified Design,  $d_3$  = New Design
  - $e_1$  = Aluminium,  $e_2$  = Aluminium + CFRP,  $e_3$  = Exotic (e.g. Ti, SiC)
- $k$  = Built to print effort / Built to spec effort = Production / (Design + Production)

✓ **Engineering Rationale**

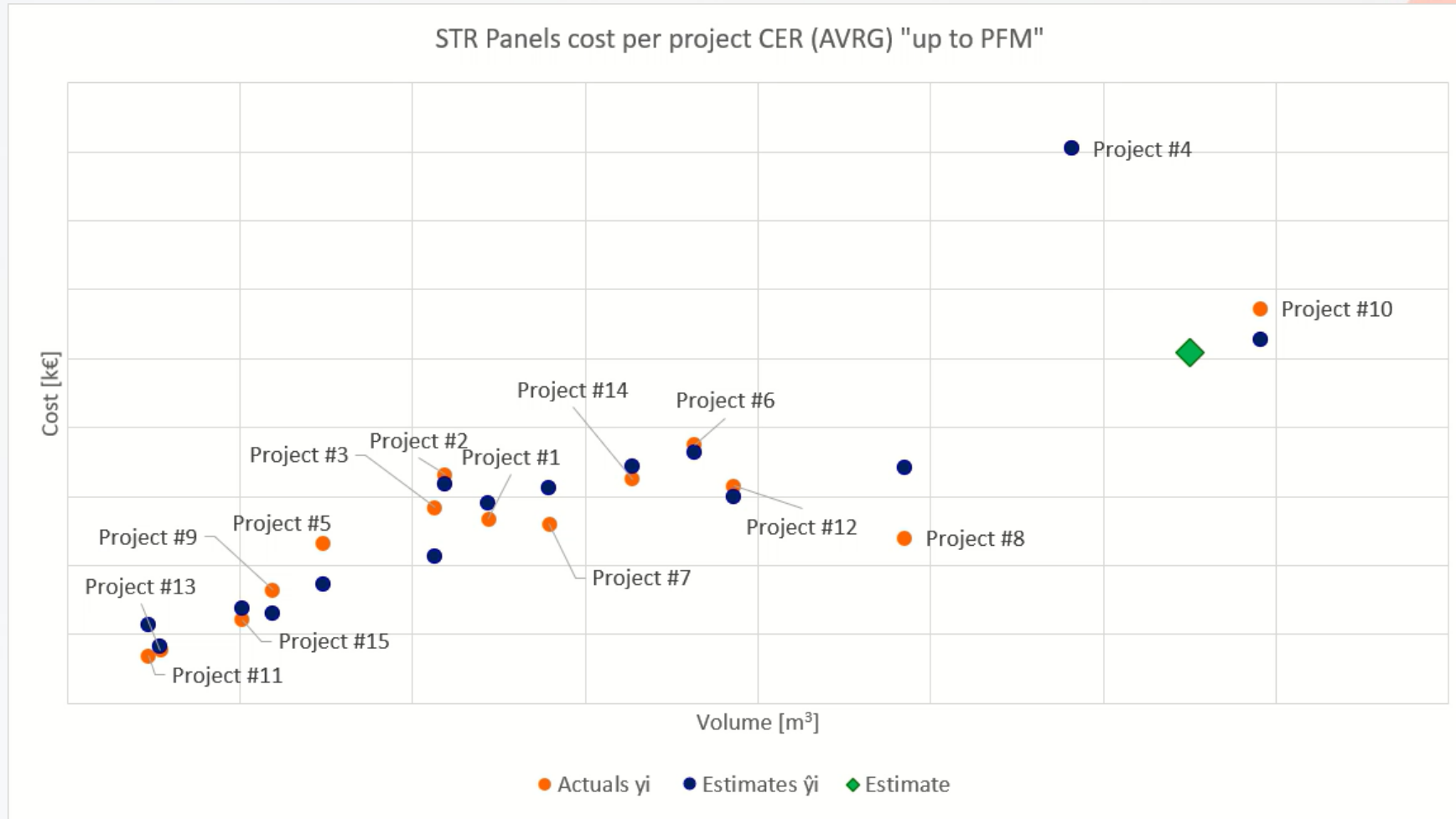
✓ **Historical Dataset**

✓ **Statistical Validation**

**Structural panel cost is driven by four key factors: how much, how mature, how complex and how much responsibility.**

# CER - Structural Panels

Results: Charts & CER quality metrics



$SSE \cdot W = 0.40$   
Weighted Prediction error

$BIAS = 0.00$   
Systematic over / under estimation

$R^2 = 0.94$   
Goodness of fit

$W = \text{Information Reliability}$

$W_{RFP} > W_{RFI}$

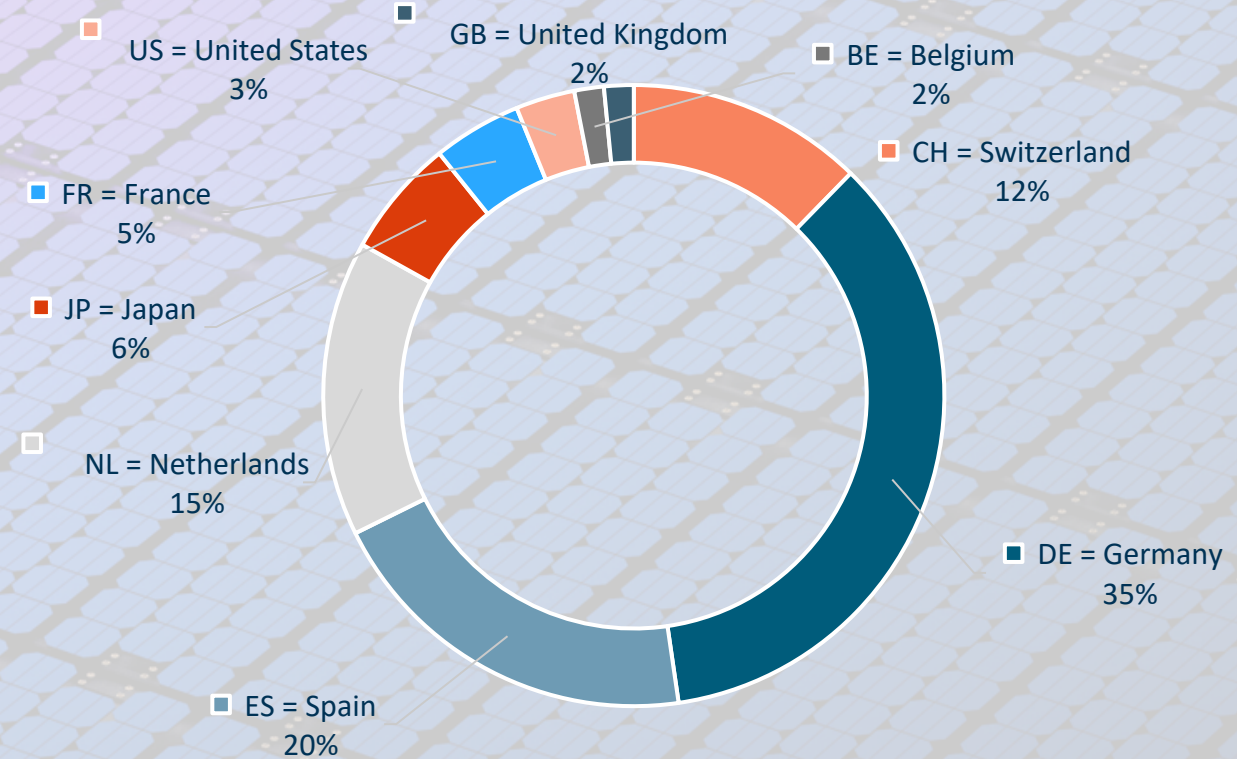
# CER – Solar Arrays

## Scope of the dataset



- ✓ Costs normalized to same Economic Conditions (EC)
- ✓ RFIs (non-binding) and RFP (binding) responses included
- ✓ Earth Observation, Science, Connectivity, Situational Awareness, Exploration programmes
- ✓ Common ESA procurement framework
- ✓ Subset of homogeneous proposals within OHb available data

## Proposals from each Country



**A diverse dataset, but a consistent procurement context**

# CER – Solar Arrays

Cost drivers' selection rationale & Formula

$$\text{Solar Array [k€]} = a + b \cdot \text{Area [m}^2\text{]}^c \cdot \text{Specific EOL Power [W/m}^2\text{]}$$



**Area**

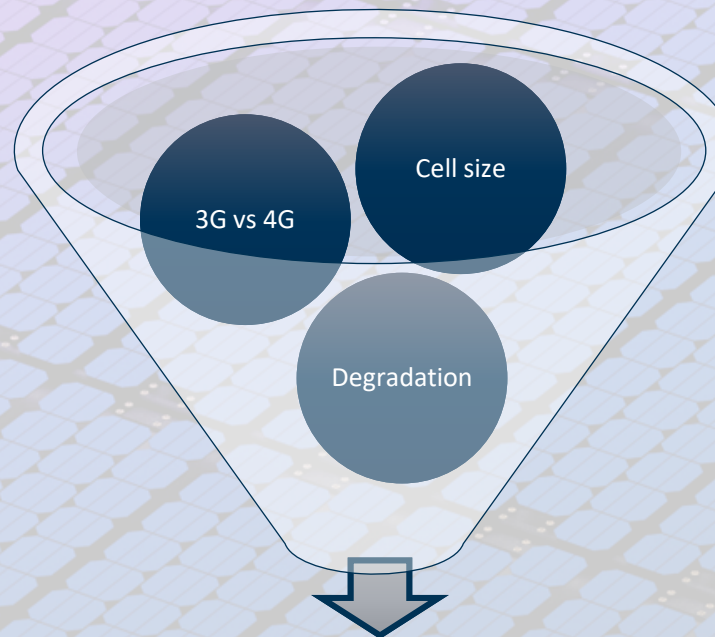
Hardware size



**Specific EOL Power**

Performance

*Candidate performance descriptors*



**Specific EOL Power**

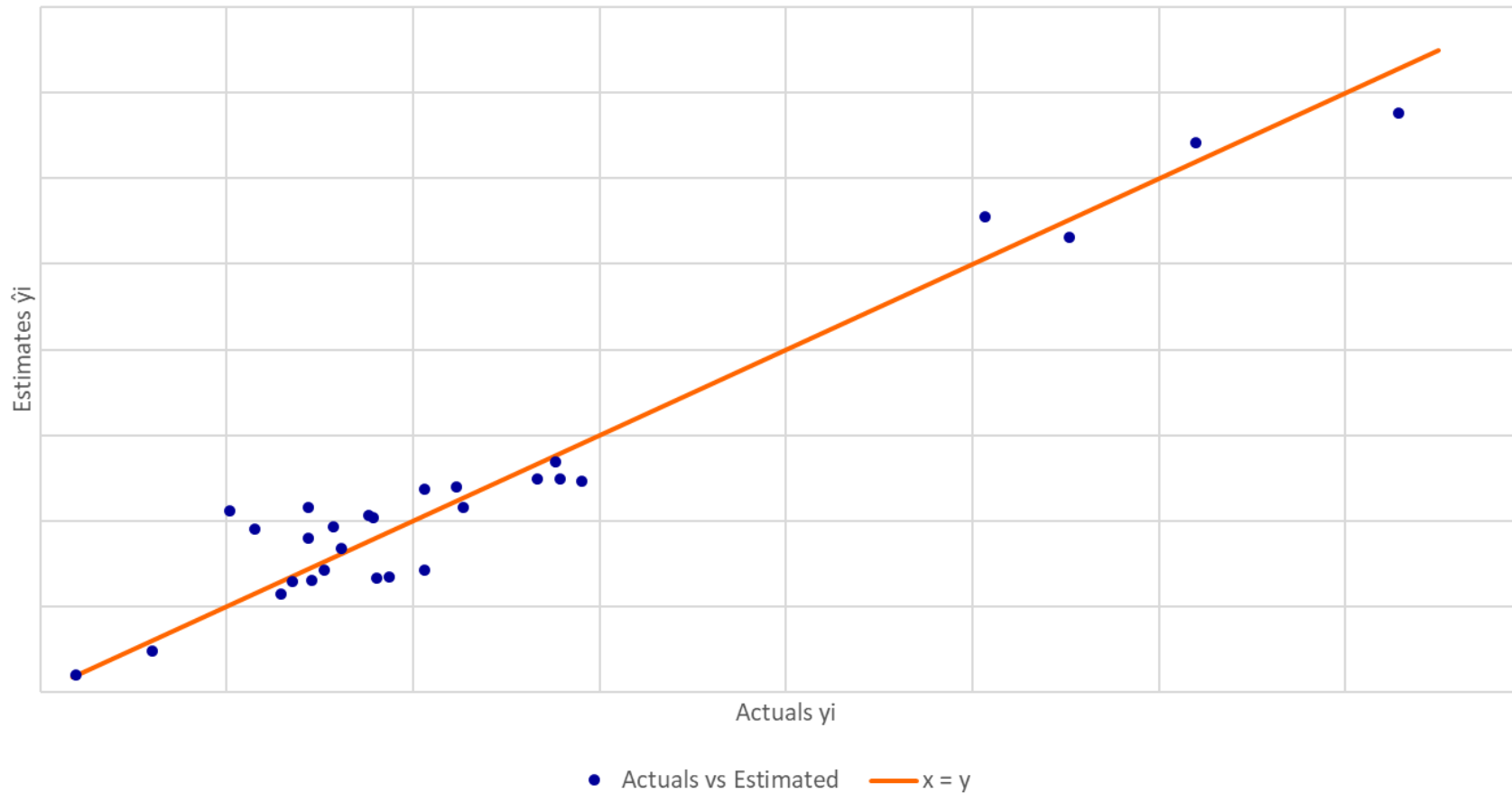
- ✓ Engineering Rationale
- ✓ Historical Dataset
- ✓ Statistical Validation

Solar-array cost depends not only on size, but also on how efficiently that area generates power.

# CER – Solar Arrays

Results: Charts & CER quality metrics

Solar Arrays - proposals dataset "up to PFM"



$$SSE \cdot W = 0.70$$

Weighted Prediction error

$$BIAS = 0.00$$

Systematic over / under estimation

$$R^2 = 0.93$$

Goodness of fit

$W = \text{Information Reliability}$

$$W_{RFP} > W_{RFI}$$

# CONCLUSION



# Conclusions

## Lessons learned & Future work

### Lessons Learned

- ◆ Engineering judgement remains essential
  - Statistical significance  $\neq$  engineering significance
- ◆ A CER is only as good as its dataset
  - Data quality drives model quality
- ◆ Simplicity often wins
  - Meaningful drivers outperform numerous drivers
- ◆ Building a CER takes time
  - But reveals insights hidden in historical data

### Future Work

- 🚀 Expand the toolset
  - Harness CER currently under development
  - Additional hardware categories to follow
- 📊 Continuous model refinement
  - Incorporate new validation datapoints
  - Periodically re-calibrate CER parameters
- ✂ Dataset pruning
  - Progressively replace older / low-maturity data

**A CER without updates is just an old opinion with a formula attached.**



# THANK YOU!

**OHV SE**

Manfred-Fuchs-Platz 2-4  
28359 Bremen  
Germany

**Phone:** +49 421 2020 8  
**Fax:** +49 421 2020 700  
**Email:** [info@ohv.de](mailto:info@ohv.de)  
**Web:** [www.ohv.de](http://www.ohv.de)